



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

ARTICLE II.

ON THE MATHEMATICAL PROBABILITY OF ACCIDENTAL LINGUISTIC RESEMBLANCES.

BY PLINY EARLE CHASE, M. A.

Read September 18, 1863.

MOST of the philological research of our day rests exclusively on a grammatical basis. The results that have been attained by such men as Edwards,* Schlegel, Bopp, and Grimm, cannot be too highly estimated, but their labors have been confessedly restricted within narrow boundaries, and they can be properly regarded only as preliminary, or introductory. The process of grouping languages into families, has already been extended nearly to its utmost practicable limits, and the question of connection between the families themselves, where no grammatical analogy can be discovered, must be solved, if it can be solved at all, by a comparison of radical words or syllables. Such comparisons have fallen into some disrepute for four principal reasons, viz.:

1. The progress of philology has brought to light old national affinities that had been previously unsuspected, showing that instances of supposed dialectic parentage were merely relations of fraternity or consanguinity, and many of the verbal derivations of the early etymologists have been consequently discarded.

2. Radical philology is a new science, and like most new sciences it has sometimes suffered from hasty generalizations, which, added to the occasional mistakes of enthusiasts, have brought undeserved reproach upon the whole study.

3. The natural desire to generalize results, has given rise to crude theories, some of which have been easily overturned, and others have been feebly propped by unsound or inconclusive arguments.

4. An undue importance has been sometimes attached to superficial analogies, and plausible grounds have thus been given to pretentious sciolists, for the summary dismissal of any newly discovered resemblance, by branding it as fanciful or accidental.

Let us give a brief passing glance at each of these points.

1. The old etymologies that have been discarded, have not, thereby, been rendered

* See Haven's *Archæology of the United States*, p. 55, for some remarks on the anticipation of Schlegel's idea, by Jonathan Edwards and others.

worthless. On the contrary, most of them have gained a new value, since they not only serve to strengthen the most important conclusions of comparative grammarians, but they also furnish a rich mine for the exploration of the searchers after verbal roots. Although it may be demonstrated that a given Latin word was not derived from the Greek, as was once mistakenly believed, the resemblance which led to the mistake still remains as an evidence of kindred origin, and if it be rightly studied it may help us to useful results. Although the French word *suivre* is undoubtedly more closely associated with the Latin root *see*, than with the Chinese *suy*, the latter may possibly have arisen from a similar law of verbal detrition.

2. The unfair advantage that has been taken of the confessed errors of etymologists, will undoubtedly, in time, be followed by a favorable reaction, and meanwhile it will have little influence in deterring those who are on the lookout for new discoveries. The unfairness is in itself an evidence that the science which is impugned is still in its infancy, and that its field is consequently mostly unexplored. Therefore, as soon as the fact becomes established that there is a sure basis for accurate research, the great probability of attaining satisfactory results will draw crowds of investigators.

3. Students who have devoted themselves the most earnestly to philology, have been the most thoroughly convinced that all languages exhibit bonds of connection which point to a common origin of some kind. The more profound their research, the deeper does this conviction usually become, and it is no mark of a candid spirit, for one whose acquirements are not such as to qualify him for pertinent criticism, to charge it to the fascination of a hobby, or of a pet theory.* Such a charge might merit consideration, if the conviction were solitary or exceptional, or if it were made the main support of a questionable hypothesis. But its approach to universality should be regarded as sufficient evidence that it is well grounded, while the different reasons that have been imagined by different investigators to explain the connection, show that they have not been led astray by devotion to a system. Whether any of their theories be rejected or believed, the facts on which they were built are incontrovertible, and deserving of careful scientific investigation.

* I am aware that some distinguished scholars, like M. Renan, deny the existence of any traces of linguistic unity, but such a denial can have little weight with any one who is at all familiar with the profound researches of the German philologists. The *a priori* probability that there was a primitive significance to every syllable, and even to every sound, and that traces of such significance are still to be found in the languages which Max Müller has classed as "Turanian," is greatly strengthened by a comparison of such dialects as the Chinese, Egyptian, and Yoruban, and it seems to me much more natural and reasonable to regard identical roots as evidences of family identity, than to attempt to explain their existence in any other way.

4. The objection to fanciful or accidental resemblances, while it is the one most frequently urged, is also the most plausible, and I have thought it worthy of a brief examination, which has led to conclusions surprising even to myself, well prepared as I was, by twenty-five years' study of radical philology, to admit the truths of which those results are the natural indications. I had supposed that an argument on which so much stress is laid, one which is urged so often and so triumphantly, must have some claims to respect, but I find that a purely and strictly accidental coincidence* in sound and sense, is nearly, if not altogether, impossible.

For, in the first place, we must reject from the category of casual resemblance, all words that are determined by fixed and known laws; such, for example, as indicate nothing more than a uniformity of idea or of vocalization, that is dependent on a uniformity of human nature in its mental and physical constitution, and all words which are clearly or even probably derived from an onomatopœia. Words of the latter class are probably not so numerous as is often supposed, and even those which may possibly have had an onomatopœic source (as *e. g.* Eng., *beef*; Fr., *bœuf*; Lat. *bov*; Gr., *βov*; Chin., *moo*), are often so transformed that their origin can only be made evident by a series of connected derivations, which serve to strengthen a conviction in the original family union of nations that are now most widely separated.

Let us then suppose that in two languages which we wish to make the subjects of comparison, all words of a merely imitative character, and all other analogues that could be reasonably referred to uniformity of organization, have been discarded. Suppose, moreover, that for other quasi-accidental and unknown reasons, there still remains a degree of resemblance so incredible that there is a precise correspondence between all the ideas represented in the two languages, and a correspondence equally precise in the sound of two-thirds of the words that they employ. Still, the word that denotes *boy* in one language, will be more likely to denote *horse*, or anything else rather than *boy*, in the other, and a single instance of merely fortuitous coincidence will be improbable. If the resemblance is still more striking, so that together with the correspondence of ideas, there is a precise agreement in the sound of *all* the syllables and words, there will probably be a single coincidence, and no more.

Some of the results of the following discussion have been anticipated by Dr. Thomas Young (Phil. Trans., 1819, p. 79, sqq.), but the importance of the subject seemed sufficient to warrant a more general investigation.

* In the present paper, I use the term *coincidence*, merely to denote words in two different languages which agree both in sound and meaning.

We will first examine the simplest case, in which there are m identical words, and m identical meanings in the two languages.

Suppose the words are similarly arranged in each language. Then, among the possible permutations of the meanings, there is only one in which they will all coincide, and none in which $m-1$ will coincide.

If all but two coincide, those two must change places. There are, therefore, as many such arrangements as we can make selections of 2 out of m , or, $\frac{m(m-1)}{2}$.

If all but three coincide, those three must change places in such a way that neither will occupy its original place. This can be done in two ways with each group of 3, and there are, therefore, twice as many such arrangements as the number of selections that we can make of 3 out of m , or, $2 \times \frac{m(m-1)(m-2)}{3!}$

Tabulating and differencing these results, it will be seen that the number of arrangements in which there are n displacements out of m , is $\Delta^n 0! \times \frac{m \dots (m-n+1)}{n!}$

The value of this expression can be readily ascertained, for

$$\begin{aligned} \Delta^n 0! &= \frac{n!}{0!} - \frac{n!}{1!} + \frac{n!}{2!} - \dots \\ n\Delta^{n-1} 0! &= \frac{n!}{0!} - \frac{n!}{1!} + \frac{n!}{2!} - \dots (-)^{n+1} \cdot 1 \quad \text{Hence, by subtraction,} \\ \Delta^n 0! &= n\Delta^{n-1} 0! (-)^n \cdot 1 \end{aligned}$$

There will \therefore be m displacements, or 0 coincidences, in $\Delta^m 0! = \frac{m!}{e} = .36788 m!$ arrangements, leaving for the probability of one or more coincidences, $.63212 m!$ out of $m!$ possible arrangements,—which is a probability of about $\frac{2}{3}$. There will, therefore, probably be at least one coincidence.

There will also be $m-1$ displacements, or 1 coincidence, in $.36788 m!$ arrangements. Deducting this amount from $.63212 m!$ there remains only a chance of $\frac{.26424}{1.00000}$ for more than one coincidence. If the words, and also the meanings, were identical in two languages, there would \therefore probably be 1 accidental coincidence, and no more, as stated above.

In order to investigate the other supposed case, it will be necessary to consider the more general expression, $\Delta^n m!$, the value of which can be conveniently ascertained in two ways. For,

$$\begin{aligned} \Delta^n m! &= \Delta^n (1 + \Delta)^m 0! \\ &= \left\{ \frac{\Delta^n}{0!} + \frac{m}{1!} \Delta^{n+1} + \frac{m(m-1)}{2!} \Delta^{n+2} + \dots \Delta^{m+n} \right\} 0! \end{aligned} \quad (1)$$

$$\text{or, } \Delta^0 m! = m! \quad \text{A}$$

$$\Delta^1 m! = m.m! \quad \text{A'}$$

$$\Delta^2 m! = m \text{ A' } + \text{A} + \text{A'}$$

$$= (m+1) \text{ A' } + \text{A} \quad \text{A''}$$

$$\begin{aligned}
\Delta^3 m! &= (m+1) \Delta'' + \Delta' + \Delta'' + \Delta' \\
&= (m+2) \Delta'' + 2 \Delta' \\
&\vdots \qquad \qquad \vdots \\
\Delta^n m! &= (m+n-1) \Delta^{n-1} m! + (n-1) \Delta^{n-2} m! \qquad (2)
\end{aligned}$$

If we compare two languages in which there are m identical meanings, and only $m-n$ identical words, there will evidently be no arrangements that admit of $m, m-1, \dots, m-n+1$ coincidences.

The $m-n$ words will retain their position in $n!$ arrangements.

There will be $m-n-1$ coincidences in $\Delta n! \times (m-n)$ arrangements. And, generally, there will be $\frac{(m-n) \dots (m-n-r+1)}{r!} \Delta^r n!$ arrangements, which admit $n+r$ displacements, or $\Delta^{m-n} n!$ arrangements in which there will be m displacements, and 0 coincidences.

In order to ascertain when the chance of a coincidence is less than $\frac{1}{2}$, and therefore cease to be a probability, we have

$$\begin{aligned}
\Delta^{m-n} n! &> \frac{e \Delta^m 0!}{2}; \text{ or, by (1)} \\
&> \Delta^{m-n} 0! + \frac{n}{0!} \Delta^{m-n+1} 0! + \dots + \frac{n}{0!} \Delta^{m-1} 0! + \Delta^m 0!
\end{aligned}$$

The ratio of n to m that will satisfy this inequality, may be found by making $\frac{m!}{2} = \Delta^m 0! + n \Delta^{m-1} 0! + \dots$ &c. Assuming 10 as a convenient value for m , extending the right hand number to four or five terms, and solving the equation, we obtain the corresponding value, $n=3.185$; or, as the ratio is independent of any particular values, $n=.3185 m$. $m-n=.6815 m$. \therefore If the number of identical words is less than .6815 of the entire number of words in each language, any accidental coincidence would be improbable. Q. E. D.*

The likelihood is not changed by any possible multiplication of the number of distinct ideas that a language may contain. For, if there are mn separate notions, to be represented by n words, while each word will have an average of m meanings, the probability of any single meaning being assigned to any particular word, is $\frac{m}{mn}$. This is precisely equivalent to $\frac{1}{n}$, which is the like probability when the number of notions is only n .

On the other hand, an increase in the number of meanings manifoldly increases the

* As a case in point, I would refer to the Cherokee alphabet. Its inventor, Sequoia, had seen our alphabet, but was ignorant of the phonetic value of any of the characters. He modified the forms of F, O, Q, U, and V, and employed, without any alteration, all the other letters except N and X, using each letter to represent a syllable. Only one of his characters has a value at all resembling our own; the letter L, which stands for the syllable TLE. This half-coincidence admirably exemplifies the calculated probability that there would be one coincidence and no more, and the equally divided chance between a single coincidence and no coincidence. The alphabet in question is given in *Schoolcraft's History of the Indian Tribes*, vol. 2, p. 228.

probability of a common origin, when the words compared present a coincidence in more than one of their meanings. For, if the concurrent meanings are apparently derived from a single primitive, we have the case which has been long and generally recognized as strengthening the evidence of family connection,—the case of parallelism in thought, which comes next in point of importance to parallelism in grammatical forms; while, if the meanings are radically distinct, the probability of each successive coincidence is diminished in a geometrical ratio, the chance of a second being only $\frac{1}{n^2}$,—and of m simultaneous coincidences, only $\frac{1}{n^m}$. Therefore, although it is desirable to ascertain the primitive or radical signification of words, before instituting a comparison, it is not absolutely essential, neither is it even so important as it is often thought.

But what can be said of synonymes? Is it not probable that when the number of verbal equivalents becomes large, the number of accidental resemblances will be proportionally large? Here, if anywhere, is the stronghold of the believers in casual similitudes, but even here their position may be easily assailed.

Granting, as before, for the influence of known and unknown laws, that two given languages represent the same ideas, and are also homonymous, the meanings being allotted to the several words at random, let us further suppose that each idea is attached to m synonymes in each language, the entire number of words being n . If the synonymes of a given idea in one language are $\alpha, \beta, \gamma, \dots \mu$, we shall still have no probable grounds to expect that either of these words will represent the same idea in the other. For the chance of the given idea being represented by any single word in the other language being $\frac{1}{n}$, the chance of its being represented by some one of the m homonymous words is $\frac{m}{n}$, which does not become a probability until $m > \frac{n}{2}$, a degree of synonymy that is wholly incredible.

Finally, if the m words in the first language are not only synonymes of a single idea, but perfect equivalents that may be taken indiscriminately, each of the words $a, b, c, \dots m$, being defined by the same set of ideas, $\alpha, \beta, \gamma, \dots \mu$, the chance of a coincidence between some one of the m ideas and either of the m homonyms in the other language, would be $\frac{m^2}{n}$, which becomes a probability when $m > (\frac{n}{2})^{\frac{1}{2}}$.* But even in that case, the chance of a second coincidence on the same word, would be only $\frac{m}{n} \times \frac{m^2}{n}$, which is less than a probability.

That the change of form which the words of every living language are constantly undergoing, has no effect on the probability of accidental resemblance, is evident from the fact that all our reasonings have been of the most general character, and they may be applied to any languages whatever, without regard to their family relationship, their present

* There would still hardly be a likelihood of more than one coincidence among all the homonyms. For the chance of such a coincidence would be indicated by $m \times \frac{m^2}{n}$, which does not become a probability until $m > (n)^{\frac{1}{3}}$.

or past condition, their relative antiquity, or any other incidental circumstances. Whenever, therefore, in the course of our linguistic comparisons, we discover any marked similarity both in sound and sense, it may safely be assumed that the resemblance is not accidental, but that it results from the operation of some adequate cause. Although, in many cases, that cause cannot be positively ascertained, we may often satisfy ourselves as to its probable character.

For example, let the subject of comparison be the Chinese Mandarin root *ma.g*, or *mi.g*, which denotes “great, vast, confused, mixed,” and other similar meanings. As analogues, we find in Sanscrit *ma.h*, to grow; *mah*, to honor; *ma’h*, to measure; *maef*, to collect, to fill, to mix; with the derivatives *mahat*, great, mighty, &c.; in Greek, *μάγανον*, *μάρος*, *μάζαρ*, *μαζρός*, *μάλα*, *μέγας*, *μίγνυμι*, *μιξ*; in Latin, *magister*, *magnus*, *misc*, *mix*, *majestas*, *mango*; in Gothic, *mickels*; Ang. Sax., *maegn*, *micel*, *mucel*; Swedish, *mycken*; Scotch, *mekyl*, *muckle*, *myche*; Spanish, *mucho*; English, *mingle*, *mongrel*, *mix*, *much*; in old Egyptian, *mah*, to fill; *mak*, to rule. The common ancestry of the Sanscrit, Greek, Latin, and Gothic dialects, is now generally admitted, while the affinity of the Chinese is still a mooted question. The chance for a merely fortuitous resemblance in the Chinese radical, may be determined in the following manner.

There are in the Mandarin dialect, eighteen initial sounds, seven which may be either medial or final, and only four which can be used as final in connection with a medial. The respective chances of accidental concurrence on the several sounds are, therefore, $\frac{1}{18}$, $\frac{1}{7}$, and $\frac{1}{4}$, and the chance that the concurrent sounds should be similarly arranged, is $\frac{1}{3!} = \frac{1}{6}$. The chance of the entire coincidence is only $\frac{1}{18} \times \frac{1}{7} \times \frac{1}{4} \times \frac{1}{6} = \frac{1}{3024}$, and it is, therefore, morally certain that the resemblance is not accidental.

The efficient causes which are most frequently set forth to explain such resemblances, are those which have already been intimated, viz.: 1. Uniformity of physical organization; 2. Uniformity of mental action; 3. Imitative nomenclature, or onomatopœia; 4. Affiliation of languages. The first may be deemed sufficient to account for a resemblance in the elementary sounds, and the second for a similarity of ideas, but neither separately nor in combination is it easy to conceive of their determining the assignment of special sounds to the expression of special ideas.

There is no plausible onomatopoetic explanation for any portion of the word except the broad vowel *a*, which, if we set aside as being thus sufficiently accounted for, the chance of casual coincidence is increased to $\frac{1}{3024} \div \frac{1}{4} = \frac{1}{756}$. Although it is possible, and perhaps even probable, that the other sounds may also have been primitively onomatopoetic, we have no right to take it for granted that they were so; and even if they were, the fact would neither diminish any probability of a common origin that might be advocated on other grounds, nor would it explain the precise arrangement of sounds which has been adopted.

There would still remain an odds of five to one against that particular arrangement, which might be readily and most satisfactorily overcome, by supposing that the root in question had been handed down to each language from some older and extinct dialect. That this is the true solution of the problem, is rendered still more likely by the following considerations.

1. Although the Chinese is entirely destitute of grammatical inflections, and it is, therefore, impossible to subject it to any grammatical comparisons, except such as are merely syntactical, its syntax, as Chev. Bunsen has demonstrated, is clearly of the Aryan type.

2. The Mandarin dialect has no final gutturals, but the terminal *g* is often replaced by a guttural in the other dialects. Thus, the mandarin word *pa.g*, to bind, to tie, becomes *pak* in Hok-Këen, and is thus naturally associated with Sanscrit *paḡ*, *paḥ*, Greek, *παῖγ*—, *παζ*—*πῆγγυ*—, Latin, *pang*, *pac*, *pax*, Ger., Eng., and Sw., *pack*, in the same way that *ma.g* is connected with a *μαζ*—, *μεγ*—, *may*—, *magn*—, &c.* Hok-Këen being in the southeast quarter of the Chinese empire, closer resemblance to the Aryan forms would naturally be expected, and the discovery of that resemblance greatly strengthens the probability of a common origin.

3. The broad vowels are often degraded in Chinese, as well as in the Aryan languages. E. g., *pa.g*, *ma.g*, become respectively *pi.g*, *mi.g*, without losing their primitive meaning or producing any change except a greater specialty of application. So in Latin, *pang-o*, *imping-o*; English, *mong-rel*, *ming-le*.

4. The Mandarin root *ma*, to add to, to increase, which may either be the primitive from which *ma.g* was derived, or an abbreviated form of the latter word, is found in *μάλα*. The double sense of greatness and seniority, is common both to Latin *major*, and Chinese *ma.g*.

From the cumulative evidence that I have thus briefly presented, I am unable to draw any other conclusion than that the Chinese and Aryan words here given, have a common origin. And finding, as I do, that comparisons equally striking, can be made with nearly every Chinese root, I can readily believe, with Bunsen and others, that the Chinese is the oldest language of which we have any record, “the monument of antediluvian speech.”

I have purposely based my calculations upon assumptions the most favorable that I could imagine, for the production of abnormal resemblances. Every deviation in any of the hypothetical postulates, such as new evidence of pre-historical national intercourse, an increase in the number of similar roots which are traceable in different families of languages, well-sustained philosophical generalizations which lead to just methods of philological study, the discovery of a new law of verbal derivation or transformation, while it may have some tendency to complicate the problem, has a still greater tendency to render its

* The Hok-Këen word *mi.g*, an incantation, similarly serves to connect the root *ma.g*, with *μάγγ-ωνον*, *magus*, &c.

further investigation unnecessary, because it diminishes, in a rapidly increasing ratio, the chances for merely accidental analogies.

We may, therefore, safely assume that any single coincidence in words of two or more syllables, or any two coincidences in radical syllables, furnish almost irresistible evidence of a national intercourse to which the coincidence is attributable. The presumption is greatly increased by every additional coincidence of either kind, and if the concurrence is frequent, or if it extends alike to derivatives and primitives, the hypothesis of mere national intercourse, by commerce or conquest, becomes less satisfactory, and it is difficult to imagine any sufficient explanation other than a common genealogy. As there are no two known languages which are destitute of a large number of such coincidences, the *à posteriori* evidence of a unity of language appears to be stronger than that of a unity of race.

Any considerable number of accidental coincidences being, as we have shown, not only improbable but morally impossible, the reasonable effort of a just philosophy, whenever they occur, is to search for some law by which they can be satisfactorily explained. By such a search, carefully conducted, philology may gradually be placed on a "positive" basis. It is certain that no adequate explanation has yet been offered for the various linguistic phenomena, except that which is based on common descent, and if none other should be found hereafter, the believers in the specific unity of man, whether upon scriptural, historical, psychological, or mere philosophical grounds, may well be gratified at such strong confirmation of their views.